CLAIMS

What is claimed is:

1. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

ordering messages on a multicast tree; and

performing aggregation of ordering primitives across said tree to minimize control traffic among nodes.

2. A method as recited in claim 1, wherein said ordering is performed on a mirror copy of an underlying shared multicast tree.

3. A method as recited in claim 1, wherein ordering of messages from rapidly changing sources, for overlapping receiver groups, and for anonymous hosts, is supported.

4. A method as recited in claim 1, further comprising distributing said ordering across nodes within the network.

5. A method as recited in claim 1, further comprising:

utilizing address extensions assigned to hosts for self-routing of messages and dynamic distribution of ordering processing load;

wherein total ordering of messages for anonymous and overlapping receiver groups in shared trees is supported.

6. A method as recited in claim 1, further comprising:

ordering messages in a diffusing computation;

wherein said messages are ordered on corresponding delivery paths from sources to receivers; and

wherein each node is responsive only to its parent and child nodes.

7. A method as recited in claim 1, further comprising:

multicasting a message from a source to a receiver set;

sending ordering information for the message to a common node on a tree elected as an ordering node for said receiver set.

8. A method as recited in claim 7, wherein said ordering information is selected from the group consisting essentially of sequence numbers and time-stamps,

9. A method recited in claim 1, wherein an ordering node sequences messages assigned to said ordering node and multicasts binding sequence numbers for final delivery to a receiver set where pending messages are to be delivered.

10. A method as recited in claim 1:

wherein a node maintains first and second message windows for ordering of multicast messages;

wherein said first window is for unordered messages which have been received but whose delivery is pending; and

wherein said second window is for messages which are correctly ordered and can be delivered to local processes.

11. A method as recited in claim 1:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

12. A method as recited in claim 1:

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node whose label is the longest common prefix among all node labels in the receiver set.

13. A method as recited in claim 1:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

14. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

ordering messages on a multicast tree in a diffusing computation;

wherein said messages are ordered on corresponding delivery paths from sources to receivers; and

wherein each node is responsive only to its parent and child nodes in said tree.

15. A method as recited in claim 14, further comprising performing aggregation of ordering primitives across said tree to minimize control traffic among nodes.

16. A method as recited in claim 14, wherein said ordering is performed on a mirror copy of an underlying shared multicast tree.

17. A method as recited in claim 14, wherein ordering of messages from rapidly changing sources, for overlapping receiver groups, and for anonymous hosts, is supported.

18. A method as recited in claim 14, further comprising distributing said ordering across nodes within the network.

19. A method as recited in claim 14, further comprising:

utilizing address extensions assigned to hosts for self-routing of messages and dynamic distribution of ordering processing load;

wherein total ordering of messages for anonymous and overlapping receiver groups in shared trees is supported.

20. A method as recited in claim 14, further comprising:

multicasting a message from a source to a receiver set;

sending ordering information for the message to a common node on a tree elected as an ordering node for said receiver set.

21. A method as recited in claim 20, wherein said ordering information is selected from the group consisting essentially of sequence numbers and time-stamps,

22. A method recited in claim 14, wherein an ordering node sequences messages assigned to said ordering node and multicasts binding sequence numbers for final delivery to a receiver set where pending messages are to be delivered.

23. A method as recited in claim 14:

wherein a node maintains first and second message windows for ordering of multicast messages;

wherein said first window is for unordered messages which have been received

but whose delivery is pending; and

wherein said second window is for messages which are correctly ordered and can be delivered to local processes.

24. A method as recited in claim 14:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

25. A method as recited in claim 14:

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node whose label is the longest common prefix among all node labels in the receiver set.

26. A method as recited in claim 14:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

27. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

ordering messages on a multicast tree;

multicasting a message from a source to a receiver set; and

sending ordering information for the message to a common node on a tree elected as an ordering node for said receiver set.

28. A method as recited in claim 27, wherein said ordering information is selected from the group consisting essentially of sequence numbers and time-stamps,

29. A method as recited in claim 27, further comprising performing aggregation of ordering primitives across said tree to minimize control traffic among nodes.

30. A method as recited in claim 27, wherein said ordering is performed on a mirror copy of an underlying shared multicast tree.

31. A method as recited in claim 27, wherein ordering of messages from rapidly changing sources, for overlapping receiver groups, and for anonymous hosts, is supported.

32. A method as recited in claim 27, further comprising distributing said ordering across nodes within the network.

33. A method as recited in claim 27, further comprising:

utilizing address extensions assigned to hosts for self-routing of messages and

dynamic distribution of ordering processing load;

wherein total ordering of messages for anonymous and overlapping receiver groups in shared trees is supported.

34. A method as recited in claim 27, further comprising:

ordering messages in a diffusing computation;

wherein said messages are ordered on corresponding delivery paths from sources to receivers; and

wherein each node is responsive only to its parent and child nodes.

35. A method recited in claim 27, wherein an ordering node sequences messages assigned to said ordering node and multicasts binding sequence numbers for final delivery to a receiver set where pending messages are to be delivered.

36. A method as recited in claim 27:

wherein a node maintains first and second message windows for ordering of multicast messages;

wherein said first window is for unordered messages which have been received but whose delivery is pending; and

wherein said second window is for messages which are correctly ordered and can be delivered to local processes.

37. A method as recited in claim 27:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

38. A method as recited in claim 27:

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node whose label is the longest common prefix among all node labels in the receiver set.

39. A method as recited in claim 27:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

40. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

multicasting a message from a source node to a receiver group;

unicasting a control message from a source node across a primary node to an ordering node for a designated multicast group or transmission, wherein said primary node aggregates messages from their subtrees and hence staggers the ordering process upward within the tree;

determining a binding sequence number for this message and a multicast to the receiver group; and

delivering messages at end hosts according to agreed-upon sequence numbers.

41. A method as recited in claim 40:

wherein said messages are delivered in an order agreed-upon by all hosts.

42. A method as recited in claim 40:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

43. A method as recited in claim 40:

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node having label that is the longest common prefix among all node labels in the receiver set.

44. A method as recited in claim 43:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

45. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

multicasting a message from a source node to a receiver group;

unicasting a control message from a source node across a primary node to an ordering node for a designated multicast group or transmission, wherein said primary node aggregates messages from their subtrees and hence staggers the ordering process upward within the tree;

determining a binding sequence number for this message and a multicast to the receiver group; and

delivering messages at end hosts according to agreed-upon sequence numbers;

wherein said messages are delivered in an order agreed-upon by all hosts.

46. A method as recited in claim 45:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

47. A method as recited in claim 45:

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node having label that is the longest common prefix among all node labels in the receiver set.

48. A method as recited in claim 47:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

49. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

multicasting a message from a source node to a receiver group;

unicasting a control message from a source node across a primary node to an ordering node for a designated multicast group or transmission, wherein said primary node aggregates messages from their subtrees and hence staggers the ordering process upward within the tree;

determining a binding sequence number for this message and a multicast to the receiver group; and

delivering messages at end hosts according to agreed-upon sequence numbers;

wherein said messages are delivered in an order agreed-upon by all hosts.

50. A method as recited in claim 49:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

51. A method as recited in claim 49:

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node having label that is the longest common prefix among all node labels in the receiver set.

52. A method as recited in claim 51:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

53. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

multicasting a message from a source node to a receiver group;

unicasting a control message from a source node across a primary node to an ordering node for a designated multicast group or transmission, wherein said primary node aggregates messages from their subtrees and hence staggers the ordering process upward within the tree;

determining a binding sequence number for this message and a multicast to the receiver group;

delivering messages at end hosts according to agreed-upon sequence numbers;

wherein said messages are delivered in an order agreed-upon by all hosts; and

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node having label that is the longest common prefix among all node labels in the receiver set.

54. A method as recited in claim 53:

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.

55. A method as recited in claim 53:

wherein each node $i$ in an acknowledgment-tree is labeled with a unique label $l(i)$, which is the prefix of all children of $i$.

56. A concurrent, multicast communication method for transmitting data packets over a network of interconnected nodes, comprising:

multicasting a message from a source node to a receiver group;

unicasting a control message from a source node across a primary node to an ordering node for a designated multicast group or transmission, wherein said primary node aggregates messages from their subtrees and hence staggers the ordering process upward within the tree;

determining a binding sequence number for this message and a multicast to the receiver group;

delivering messages at end hosts according to agreed-upon sequence numbers;

wherein said messages are delivered in an order agreed-upon by all hosts;

wherein, for each set of messages destined to a particular multicast group, or set of hosts, an ordering node is elected by virtue of being the node having label that is the longest common prefix among all node labels in the receiver set; and

wherein each ordering node gathers sequence number bids set *en route* by primary nodes deciding on a globally valid number, and multicasts the respective message to the receiver set with a final and binding sequence number directive.